The past decade and future of AI's Impact on Society[1]

Joanna J. Bryson

University of Bath, Department of Computer Science

Abstract

Artificial intelligence (AI) is a technical term referring to artefacts used to detect contexts or to effect actions in response to detected contexts. Our capacity to build such artefacts has been increasing, and with it the impact they have on our society. This chapter first documents the social and economic changes brought about by our use of AI, particularly but not exclusively focussing on the decade since the 2007 advent of smart phones, which contribute substantially to 'big data' and therefore the efficacy of machine learning. It then projects from this political, economic, and personal challenges confronting humanity in the near future, including policy recommendations. Overall, AI is not as unusual a technology as expected, but this very lack of expected form may have exposed us to a significantly increased urgency concerning familiar challenges. In particular, the identity and autonomy of both individuals and nations is challenged by the increased accessibility of knowledge.

1. Introduction

The past decade, and particularly the past few years, have been transformative for artificial intelligence (AI) not so much in terms of what we can do with this technology as what we *are* doing with it. Some place the advent of this era to 2007, with the introduction of smart phones. As I detail below, at its most essential, intelligence is just intelligence, whether artefact or

[1]An older version of some of this material was delivered to the OECD (Karine Perset) in May 2017 under the title "Current and Potential Impacts of Artificial Intelligence and Autonomous Systems on Society", and contributes to their efforts and documents of late 2018 and early 2019.

animal. It is a form of computation, and as such a transformation of information. The cornucopia of deeply personal information that resulted from the wilful tethering of a huge portion of society to the Internet has allowed us to pass immense explicit and implicit knowledge from human culture via human brains into digital form. Here we can not only use it to operate with humanlike competence, but also produce further knowledge and behaviour by means of machine-based computation.

For decades—even prior to the inception of the term—AI has aroused both fear and excitement as humanity contemplates creating machines in our image. This expectation that *intelligent* artefacts should by necessity be *humanlike* artefacts blinded most of us to the important fact that we have been achieving AI for some time. While the breakthroughs in surpassing human ability at human pursuits such as chess (Hsu, 2002), go (Silver et al., 2016), and translation (Wu et al., 2016) make headlines, AI has been a standard part of the industrial repertoire since at least the 1980s. Then production-rule or 'expert' systems became a standard technology for checking circuit boards and detecting credit card fraud (Liao, 2005). Similarly, machine learning (ML) strategies like genetic algorithms have long been used for intractable computational problems like scheduling, and neural networks not only to model and understand human learning, but also for basic industrial control and monitoring (Widrow et al., 1994). In the 1990s probabilistic and Bayesian methods revolutionised ML and opened the door to some of the most pervasive AI technologies now available: search through massive troves of data (Bishop, 1995). This search capacity included the ability to do semantic analysis of raw text, astonishingly enabling Web users to find the documents they seek out of trillions of Web pages just by typing just a few words (Lowe, 2001; Bullinaria and Levy, 2007).

This capacity to use AI for discovery has been extended not only by the massive increase of digital data and available computing power, but also by innovations in AI and ML algorithms. We are now searching photographs, videos, and audio (Barrett et al., 2016; Wu et al., 2016). We can translate, transcribe, read lips, read emotions (including lying), forge signatures and other handwriting, and forge video (Assael et al., 2016; Eyben et al., 2013; Deng et al., 2017; Haines et al., 2016; Reed et al., 2016; Vincent, 2016; Hancock et al., 2007; Chung and Zisserman, 2017; Schuller et al., 2016; Sartori et al., 2016; Thies et al., 2016). Critically, we can forge real-time audio/video during live transmissions, allowing us to choose the words millions 'witness', particularly for celebrities such as politicians for whom there is

already a great deal of data for composing accurate models (Thies et al., 2016; Suwajanakorn et al., 2017). At the time of this writing, there is increasing evidence that the outcomes of the 2016 US presidential election and UK referendum on EU membership were both altered by the AI detection and targeting of 'swing voters' via their public social media use (Cadwalladr, 2017a,b; ICO, 2018), not to mention the AI-augmented tools used in cyberhacking (Brundage et al., 2018). AI is here now, available to and benefiting us all. But its consequences for our social order are not only not understood, but have until recently barely even yet the subject of study (Bryson, 2015). Yet now too, with advances in robotics, AI is entering our physical spaces in the form of autonomous vehicles, weapons, drones, and domestic devices, including 'smart speakers' (really, microphones) and even games consoles (Jia et al., 2016). We are becoming surrounded by—even embedded in—pervasive automated perception, analysis, and increasingly action.

What have been and will be the impacts of pervasive synthetic intelligence? How can society regulate the way technology alters our lives? In this article, I begin by presenting a clean, clear set of definitions for the relevant terminology. I then review concerns and suggested remediations with respect to technology. Finally, I make expert though unproven recommendations concerning the value of individual human lives, as individual human *capacities* come increasingly under threat of redundancy to automation.

2. Definitions

The following definitions are not universally used, but derive from a well established AI text Winston (1984), as well as from the study of biological intelligence (Barrows 2000, attributed to Romanes 1883). They are selected for clarity of communication at least local to this chapter, about the existing and potential impacts of intelligence, particularly in machines. *Intelligence* is the capacity to do the right thing at the right time, in a context where doing nothing (making no change in behaviour) would be worse. Intelligence then requires:

- the capacity to perceive *contexts* for action,
- the capacity to *act*, and
- the capacity to *associate* contexts to actions.

By this definition, plants are intelligent (Trewavas, 2005). So is a thermostat (McCarthy, 1983; Touretzky, 1988). They can perceive and respond to context e.g. plants to the direction of light, thermostats to temperature. We further discriminate a system as being *cognitive* if it is able to modify its intelligence, something plants and at least mechanical thermostats cannot do. Cognitive systems are able to learn new contexts, actions, and/or associations between these. This comes closer to the conventional definition of 'intelligent'.

Intelligence as I defined it here is a strict subset of *computation*, the transformation of information. Note that computation is a physical process, it is not math. It takes time, space, and energy. Intelligence is the subset of computation that transforms a context into action.

Artificial intelligence (AI), by convention, is a term used to describe (typically digital) artefacts that extend any of the capacities related to natural intelligence. So for example, machine vision, speech recognition, pattern recognition, and fixed (unlearning) production systems are all considered examples of AI, with algorithms that can be found in standard AI textbooks (Russell and Norvig, 2009). These can also all be seen as forms of computation, even if their outputs aren't conventionally seen as action. If we embrace though the lessons of embodied robotics (see below) then we might extend this definition to include as AI *any* artefact that extends our own capacities to perceive and act. Although this would be an unusual definition, it might also give us a firmer grip on the sorts of changes AI brings to our society, by allowing us to examine a longer history of technological interventions.

Machine learning (ML) is any means of programming AI that requires not only conventional hand coding, but also a component of automated generalisation over presented data by means of accumulating statistics on that data (Murphy, 2012; Erickson et al., 2017). Often but not necessarily ML comes down to seeking regularities in data that are associated with categories of interest, including appropriate opportunities for particular actions. ML is also often used to capture associations, and can be used to acquire new action skills, for example from demonstration (Huang et al., 2016).

Note that all ML still involves a hand-programmed component. The mere conceptualisation or discovery of an algorithm never leads to a machine capable of sensing or acting springing spontaneously into existence. All AI is by definition an *artefact*, brought into being by deliberate human acts. Something must be built and designed to connect some data source to some representation before any learning can occur. All intelligent systems have

an *architecture*, a layout through which energy and information flows, and nearly always including locations where some information is retained, termed *memory*. The design of this architecture is called *systems engineering*, it is at this point that a systems safety and validity should be established. Contrary to some outrageous but distressingly frequent claims, AI safety is not a new field. Systems engineering in fact predates computers (Schlager, 1956), and has always been a principle component of computer science education. AI has long been integrated into software, as documented in the introduction, so there is a long history of it being engineered so in safe ways (e.g. Bryson, 2003; Chessell and Smith, 2013).

Robots are artefacts that sense and act in the physical world, and in real time. By this definition a smart phone is a (domestic) robot. It has not only microphones but also a variety of proprioceptive sensors that allow it to know when its orientation is changing or it is falling. Its range of actions includes intervening with its user and transmitting information including instructions to other devices. The same is true of may some game consoles and digital home assistants—'smart spakers'/microphones like Google Home, Amazon's Echo (Alexa), or Microsoft's Cortina.

Autonomy is technically the capacity to act as an individual (Armstrong and Read, 1995; Cooke, 1999). So for example a country loses its autonomy if either its institutions collapse such that only its citizens' individual actions have of efficacy, or if its institutions come under the influence of other agencies or governments to such an extent that again its own government has no impact on its course of actions. Of course, either extreme is very unusual. In fact, for social animals like humans autonomy is never absolute (Gilbert et al., 2012). Our individual intelligence determines many of our actions, but some cells may become cancerous in pursuit of their own goals contra our overall well being (Hanahan and Weinberg, 2011). Similarly, we fully expect a family, place of work, or government, to have impact on our actions. We also experience far more social influence implicitly than we are ever aware of (Devos and Banaji, 2003). Nevertheless, we are viewed as autonomous, because there is an extent to which our own individual intelligence also influences our behaviour. A technical system able to sense the world and select an action specific to its present context is therefore called 'autonomous' even though its actions will ultimately be determined by some combination of the designers that constructed its intelligence and its operators. Operators may influence AI in real time, and will necessarily influence it in advance by setting parameters of its operation, including when and where it operates, if at

all. As discussed earlier, designers call the system into existence and determine its capacities, particularly what information it has access to and what actions it can take. Even if a designer chooses to introduce an element of chance such as dependence on the present environment or a random-number generator into the control of an AI system, that inclusion is still the deliberate choice of the designer.

3. Concerns about AI and Society

AI is core to some of the most successful companies in history in terms of market capitalisation — Apple, Alphabet, Microsoft and Amazon. Along with Information and Communication Technology (ICT) more generally, AI has revolutionised the ease with which people from all over the world can access knowledge, credit, and other benefits of contemporary global society. Such access has helped lead to massive reduction of global inequality and extreme poverty, for example by allowing farmers to know fair prices, best crops, and giving them access to accurate weather predictions (Aker and Mbiti, 2010).

AI is the beneficiary of decades of regulatory policy: research and deployment has so far been largely up-regulated with massive government and other capital investment (Brundage and Bryson, 2017; Miguel and Casado, 2016; Technology Council Committee on Technology, 2016). Although much of the emphasis of later parts of this chapter focusses on possible motivations for or mechanisms of regulatory restriction on AI, it should be recognised that:

1. Any such AI policies should and basically will always be developed and implemented in the light of the importance of respecting the positive impacts of technology as well[2].
2. No one is talking about introducing regulation to AI. AI already exists in a regulatory framework (Brundage and Bryson, 2017; O'Reilly, 2017), what we are discussing is whether that framework needs optimising.

[2]See further for upside analysis the Obama administration's late AI policy documents (Technology Council Committee on Technology, 2016; of the President, 2016). For reasons of space and focus I also do not discuss here the special case of military use of AI. That is already the subject of other significant academic work, and is generally regulated through different mechanisms than commercial and domestic AI. See though Brundage et al. (2018); ICRC (2018)

3. Regulation has so far mostly been entirely constructive, with governments providing vast resources to companies and universities developing AI. Even where regulation constrains, informed and well-designed constraint can lead to more sustainable and even faster growth.

Having said this academics, technologists, and the general public have raised a number of concerns that may indicate a need for down-regulation or constraint. Smith (2018), president of Microsoft, recently asserted:

> [Intelligent[3]] technology raises issues that go to the heart of fundamental human rights protections like privacy and freedom of expression. These issues heighten responsibility for tech companies that create these products. In our view, they also call for thoughtful government regulation and for the development of norms around acceptable uses. In a democratic republic, there is no substitute for decision making by our elected representatives regarding the issues that require the balancing of public safety with the essence of our democratic freedoms.

In this section I categorise perceived risks by the sort of policy requirements they are likely to generate. I also make recommendations about whether these are non-problems, problems of ICT or technology more generally, or problems special to AI, and in each case what the remedy may be.

3.1. Artificial General Intelligence (AGI) and Superintelligence

I start with some of the most sensational claims — that as artificial intelligence increases to the point that it surpasses human abilities, it may come to take control over our resources and outcompete our species, leading to human extinction. As mentioned in Sec 1, AI is already superhuman in many domains. We can already do arithmetic better, play chess and go better, transcribe speech better, read lips better, remember more things for longer, and indeed be faster and stronger with machines than unaided. While these capacities have disrupted human lives including employment (see below), they have in no way lead to machine ambition.

Some claim that the lack of machine ambition or indeed domination is because the forms of AI generated so far are not sufficiently general. The

[3]Here, facial recognition.

term *artificial general intelligence* (AGI) is used to describe two things: AI capable of learning anything without limits, and human-like AI. These two meanings of AGI are generally conflated, but such conflation is incoherent, since in fact human intelligence has significant limitations. Understanding the limitations of human intelligence is informative because they relate also to the limits of AI.

Limitations on human intelligence derive from two causes: combinatorics and bias. The first, combinatorics, is a universal problem affecting all computation and therefore all natural and artificial intelligence. : *combinatorics* (Sipser, 2005). If an agent is capable of 100 actions, then it is capable of 10,000 2-step plans. Since humans are capable of far more than 100 different actions and perform far more than two actions even in a day, we can see that the space of possible strategies is inconceivably vast, and cannot be easily conquered by any scale of intelligence (Wolpert, 1996b).

However, computer science has demonstrated that some ways to exploring such vast spaces are more effective than others, at least for specific purposes (Wolpert, 1996a). Most relevantly to intelligence, concurrent search by many processors simultaneously can be effective provided that the problem space can be split between them, and that a solution once found can be both recognised and communicated (Grama, 2003). The reason human technology is so much more advanced than other species' is because we are far more effective at this strategy of concurrent search, due to our unique capacity to share advances or 'good tricks' via language (Dennett, 2013; Bryson, 2008, 2015; van Schaik et al., 2017). Our culture's increasing pace of change is in part due in part to the unprecedented number of individuals with good health and education connected together by ICT, but also to our augmentation of our search via machine computing. Our increasing capacities for AI and artifactual computation more generally increase further our potential rate of exploration; quantum computation could potentially accelerate these far further (Williams, 2010). However, note that these advantages do not come for free. Doing two computations at once may double the speed of the computation if the task was perfectly divisible, but it certainly doubles the amount of space and energy needed to do the computation. Quantum computing is concurrent in space as well as time, but its energy costs are so far unknown, and very likely to be exorbitant.

Much of the recent immense growth of AI has been largely due to improved capacities to 'mine' using ML the existing discoveries of humanity and nature more generally (Caliskan et al., 2017; Moeslund and Granum, 2001;

Calinon et al., 2010). The outcomes of some of our previous computation are stored in our culture, and biological evolution can also be though of as a massive parallel search, where the outcomes are collated very inefficiently, only as fast as the best genes manage to reproduce themselves. We can expect this strategy of mining past solutions to soon plateau, when artificial and human intelligence come to be sharing the same, though still-expanding boundary of extant knowledge.

The second source of limitations on human intelligence, which I called 'bias' above, are those special to our species. Given the problems of combinatorics, all species only explore a tiny subset of possible solutions, and in ML such focus is called *bias*. The exact nature of any biological intelligence is part of its evolutionary niche, and is unlikely to be shared even by other biological species except to the extent that they have similar survival requirements and strategies (Laland et al., 2000). Thus we share many of our cognitive attributes—including perception and action capacities, and importantly, motivations—with other apes. Yet we also have specialist motivations and capacities reflecting our highly social nature (Stoddart, 1990). No amount of intelligence in itself necessitates social competitiveness, neither does it demand the desire to be accepted by an ingroup, to dominate an outgroup, nor to achieve recognition within an ingroup. These are motivations that underlie human cooperation and competition that result from our evolutionary history (Mace, 1998; Lamba and Mace, 2012; Jordan et al., 2016; Bryson et al., 2017); further, they vary even among humans (Herrmann et al., 2008; Van Lange et al., 1997; Sylwester et al., 2017). For humans, social organisations easily varied to suit a politico-economic context are a significant survival mechanism (Stewart et al., 2018).

None of this is necessary—and much of it is even incoherent—from the perspective of an artefact. Artefacts are definitionally designed by human intent, not directly by evolution. With these intentional acts of authored human creation[4] comes not only human responsibility, but an entirely different landscape of potential rewards and design constraints (Bryson et al., 2017; Bryson, 2018).

Given all of the above, AGI is obviously a myth — in fact, two orthogonal

[4]The choice to create life through childbirth is not the same. While we may author some of childrearing, the dispositions just discussed are shared with other primates, and are not options left to parents to authors.

myths.

1. No amount of natural or artificial intelligence will be able to solve all problems, and
2. even extremely powerful AI is exceedingly unlikely to be very human-like, because it will embody an entirely different set of motivations and reward functions.

These assertions however do not protect us from another, related concern. *Superintelligence* is a term used to describe the situation when a cognitive system not only learns, but learns how to learn. Here again there are two component issues. First, at this point an intelligence should be able to rapidly snowball to such an extent that it would be incomprehensible to ordinary human examination. Second, even if the intelligence was carefully designed to have goals aligned with human needs, it might develop for itself unanticipated subgoals that are not. For example, a chess-playing robot might learn to shoot the people that deprive it of sufficient resources to improve its game play by switching it off at night, or a filing robot might turn the planet into paperclips in order to ensure all potential papers can be adequately ordered (Bostrom, 2012).

These two examples are ludicrous if we remember that all AI systems are designed and a matter of human responsibility. No one has ever made a chess program that represents information concerning any resources not on the chessboard (with the possible exception of time), nor with the capacity to fire a gun. The choice of capacities and components of a computer system is again part of its architecture; As I mentioned earlier, the systems engineering of architecture is an important component to extant AI safety, and as I will say below (Sec 4.3), it can also be an important means for regulating AI.

However, the concept of superintelligence itself is not ludicrous; it is clear that systems that learn to learn can and do experience exponential growth. The mistake made by futurists concerned with superintelligence is to think that this situation is only a possible future. In fact, it is an excellent description of human culture over the last 10,000 years, since the innovation of writing (Barnosky, 2008; Haberl et al., 2007). The augmentation of human intelligence with technology has indeed resulted in a system that has not been carefully designed and results in unintended consequences. Some of these consequences are very hazardous, such as global warming and the reduction of species diversity.List and Pettit (2011) make a similar point when they call human organisations such as corporations or governments 'AI'.

As I mentioned I will return to the importance of architecture and design again, but it is worth emphasising again here the necessity of such biases and limits. Robots make it particularly apparent that behaviour depends not only on computational capacities but also on other system attributes such as physical capacities. Digital manipulation such as typing or playing the flute is just not an option for either a smart phone or a snake, however intelligent. Motivations are similar. Unless we design a system to have anthropomorphic goals, social perception, and social behaviour capacities, we are not going to see it learning to produce anthropomorphic social behaviour like seeking to dominate conversations, corporations, or countries. If corporations do show these characteristics, it is because of the expression of the human components of their organisation, and also because of the undisciplined, evolutionary means by which they accrete size and power. From this example we can see that it is possible for an AI system—at the very least by the List and Pettit (2011) argument, to express superintelligence, which implies that such intelligent systems should be regulated to avoid this.

From the above I conclude that the problem of superintelligence is real but not special to AI; it is rather one our cultures already face. AI is however now a contributing factor to our capacity to excel, but this may also lead us to learn to better self regulate—that is, govern—as it has several times in the past (Milanovic, 2016; Scheidel, 2017). Even were AGI to be true and the biological metaphor of AI competing by natural selection to be sound, there is no real reason to believe that we would be extinguished by AI. We have not extinguished the many species (particularly of microbial) on which we ourselves directly depend. Considering unintended consequences of the exponentially increasing intelligence of our entire socio-technical system (rather than AI on its own) does however lead us to more substantial concerns.

3.2. Inequality and employment

For centuries there have been significant concerns about the displacement of workers by technology (Autor, 2015). There is no question that new technologies do disrupt communities, families, and lives, but also that historically the majority of this disruption has been for the better (Pinker, 2012). In general lifespans are longer and infant mortality lower than ever before, and these indicators are good measures of contentedness in humans, as low infant mortality in particular is well associated with political stability (King and Zeng, 2001).

However some disruption does lead to political upheaval, and has been recently hypothesised to associate with the rise of AI. Income (and presumably wealth) inequality is highly correlated with political polarisation (McCarty et al., 2016). Political polarisation is defined by the inability of political parties to cooperate in democratic governance, but periods of polarisation are also characterised by increases in identity politics and political extremism. Political polarisation and income inequality covary but either can lead the other; the causal factors underlying the relationship are not well understood (Stewart et al., 2018). What is known is that the last time these measures were as high as they are now (at least in the OECD) was immediately before and after the First World War. Unfortunately, it took decades of policy innovation, a global financial crisis, and a second world war before inequality and polarisation were radically reduced and stabilised in the period 1945–1978 (Scheidel, 2017), though note that in some countries such as the USA and UK the second shock of the financial crisis was enough.

Fortunately, we now know how to redress this situation—redistribution lowers inequality. After the Second World War, when tax rates were around 50%, modern welfare states were built or finalised, transnational wealth extraction was blocked (Bullough, 2018), and both income inequality and political polarisation were kept low for over 20 years. During this time, wages also kept pace with productivity (Mishel, 2012). However, some time around 1978 wages plateaued, and both inequality and political polarisation began rising, again in the OECD[5]. The question is what caused this to happen. There are many theories, but given the starkness of the shift on many metrics it looks more like a change in policy than of technology. This could reflect geopolitical changes of the time — it could signal for example the point at which economically-influential members of the OECD detected the coming end of the Cold War, and shifted away from policies designed to combat the threat of Communist uprisings.

Regardless of the causes, with respect to AI, the fact that similar political and economic trends occurred in the late 1800s again means that this is not a special concern of any one technology. While as mentioned there is so

[5]Importantly, globally, inequality is falling, due to ICT and possibly other progress such as the effective altruism movement and data-lead philanthropy in the developing world. See earlier discussion (p. 6 and Milanovic (2016); Bhorat et al. (2016); Singer (2015); Gabriel (2017).

far no consensus on causes, in ongoing research I with other authors[6] are exploring the idea that some technologies reduce costs that had traditionally maintained diversity in the economic system. For example, when transport costs are high, one may choose to use a nearby provider rather than finding the global best provider for a particular good. Similarly, lack of information transparency or scaling capacity may result in a more diverse use of providers. Technical innovations (including in business process) may overcome these costs and allow relatively few companies to dominate. Examples from the late 19thC might include the use of oil, rail, and telegraph; the improvement of shipping and newspaper delivery.

Where a few providers receive all the business, they will also receive all of the wealth. Governance is a primary mechanism of redistribution (Landau, 2016), thus revolutions in technology may require subsequent revolutions in governance in order to reclaim equilibrium (Stewart et al., 2018). The welfare state could be one such example (Scheidel, 2017). We will return to discussing the possible need for innovations in governance below (Sec 4).

To return to AI or more likely ICT, even if these technologies are not unique in contributing to inequality and political polarisation, they may well be the principle component technologies presently doing so. Further, the public and policy attention currently directed towards AI may afford opportunities to both study and address the core causes of inequality and polarisation, particularly if AI is seen as a crisis (Tepperman, 2016). Nevertheless, it is worth visiting one hypothesised consequence of polarisation in particular. An increase in identity politics may lead to the increased use of beliefs to signal ingroup status or affiliation (Iyengar et al., 2012; Newman et al., 2014), which would unfortunately decrease their proportional use to predict or describe the world — that is, to reflect facts. Thus ironically the age of information may not universally be the age of knowledge, but rather also an age of disinformation[7].

This reliance on beliefs as ingroup indicator may influence another worrying trait about contemporary politics: loss of faith in experts. While occasionally motivated by the irresponsible use or even abuse of position by some

[6]particularly Nolan McCarty

[7]I personally suspect that some of the advanced political risk taking e.g. in election manipulation may be a result of those who fear the information age because of its consequences in terms of catching illegal financial conduct such as money laundering and fraud.

experts, in general losing access to experts' views is a disaster. The combinatorial explosion of knowledge mentioned in Sec 3.1 also means that no one, however intelligent, can master in their lifetime all human knowledge. If society ignores the stores of expertise it has built up—often through taxpayer funding of higher education—it sets itself at a considerable disadvantage.

These concerns about the nature and causes of "truthiness" in what should be the information age lead also to our next set of concerns, about the use of personal information.

3.3. Privacy, personal liberty, and autonomy

When we consider the impact of AI on individual behaviour, we now come to a place where ICT more clearly has a unique impact. There have long been periods of domestic spying which have been associated with everything from prejudiced skew in opportunities to pogroms. However, ICT is now allowing us to keep long-term records on anyone who produces storable data — for example, anyone with bills, contracts, digital devices, or a credit history, not to mention any public writing and social media use. That is, essentially, everyone.

It is not only the storage and accessibility of digital records that changes our society; it is the fact that these can be searched using algorithms for pattern recognition. We have lost the default assumption of anonymity by obscurity (Selinger and Hartzog, 2017). We are to some extent all celebrities now: any one of us can be identified by strangers, whether by facial recognition software or data mining of shopping or social-media habits (Pasquale, 2015). These may indicate not just our identity but our political or economic predispositions, and what strategies might be effective for changing these (Cadwalladr, 2017a,b). ML allows us to discover new patterns and regularities of which we may have had no previous conception. For example that word choice or even handwriting pressure on a digital stylus can indicate emotional state, including whether someone is lying (Bandyopadhyay and Hazra, 2017; Hancock et al., 2007), or a pattern of social media use predict personality categories, political preferences, and even life outcomes (Youyou et al., 2015).

Machine learning has enabled near-human and even super-human abilities in transcribing speech from voice, recognising emotions from audio or video recordings, as well as in forging handwriting or video (Valstar and Pantic, 2012; Griffin et al., 2013; Eyben et al., 2013; Kleinsmith and Bianchi-Berthouze, 2013; Hofmann et al., 2014; Haines et al., 2016; Reed et al., 2016;

Vincent, 2016; Thies et al., 2016; Deng et al., 2017). The better a model we have of what people are likely to do, the less information we need to predict what an individual will do next (Bishop, 2006; Youyou et al., 2015). This principle allows forgery by taking a model of a person's writing or voice, combining it with a stream of text, and producing a 'prediction' or transcript of how that person would likely write or say that text (Haines et al., 2016; Reed et al., 2016). The same principle might allow political strategists to identify which voters are likely to be persuaded if not to change party affiliation, at least to increase or decrease their probability of turning out to vote, and then to apply resources to persuade them to do so. Such a strategy has been alleged have impacted significant recent elections in the UK and USA (Cadwalladr, 2017a,b; ICO, 2018); if so they were almost certainly tested and deployed earlier in other elections less carefully watched.

Individuals in our society might then reasonably fear the dissemination of their actions or beliefs for two reasons: first because it makes them easier to predict and therefore manipulate, and second because it exposes them to persecution by those who do not approve of their beliefs. Such persecution could range from bullying by individuals, through to missed career or other organisational opportunities, and on to in some unstable (or at least unethical) societies, imprisonment or even death at the hands of the state. The problem with such fears is not only that the stress of bearing them is itself noxious, but also that in inhibiting personal liberty and free expression we reduce the number of ideas disseminated to society as a whole, and therefore limit our ability to innovate (Mill, 1859; Price, 1972). Responding to both opportunities and challenges requires creativity and free thinking at every level of society.

3.4. Corporate autonomy, revenue, and liability

These considerations of personal autonomy lead directly to the final set of concerns I describe here, which is not one frequently mentioned. Theoretical biology tells us that where there is greater communication, there is a higher probability of cooperation (Roughgarden et al., 2006). While cooperation is often wonderful, it can also be thought of as essentially moving some portion of autonomy from the individual to a group (Bryson, 2015). Recall from the Sec 2 definitions that the extent of autonomy an entity has is the extent to which it determines its own actions. Individual and group autonomy must to some extent trade off, though there are means of organising groups that offer more or less liberty for their constituent parts. Thus the limits on

personal liberty just described may be a very natural outcome of introducing greater capacity for communication. Here again, I again refer to all of ICT, but AI and ML with their capacity to accelerate search for both solutions and collaborators are surely a significant component, and possibly game-changing.

One irony here is that many people think that bigger data is necessarily better, but better for what? Basic statistics teaches us that the number of data points we need to make a prediction is limited by the amount of variation in that data, providing only that the data is a true random sample of its population[8]. The extent of data we need for science or medicine may require only a minuscule fraction of a population. However, if we want to spot specific individuals to be controlled, dissuaded, or even promoted, then of course we want to "know all the things."

But changing the costs and benefits of investment at the group level have more consequences than only privacy and liberty. ICT facilitates blurring the distinction between customer and corporation, or even the definition of an economic transaction. This has so far largely unrecognised though see Perzanowski and Schultz (2016); Frischmann and Selinger (2016). Customers now do real labour for the corporations to whom they give their custom: pricing and bagging groceries, punching data at ATMs for banks, filling in forms for airlines and so forth (Bryson, 2015). The value of this labour is not directly remunerated—we assume that we receive cheaper products in return, and as such our loss of agency to these corporations might be seen as a form of bartering. They are also not denominated, obscuring the value of this economy. Thus ICT facilitates a black or at least opaque market that reduces measured income and therefore tax revenue where taxation is based on denominated turnover or income. This problem holds for everyone using Internet services and interfaces, even ignoring the problematic definitions of employment raised by platforms (though see O'Reilly, 2017). Our improving capacity to derive value and power while avoiding revenue may also help explain the mystery of our supposed static productivity (Brynjolfsson et al., 2017).

This dynamic is most stark in the case of "free" Web services. Clearly we

[8]This caveat is *very* important. Much data derived from e.g. governments or commerce may well have strong biases over who is represented or even how well that data is transcribed. Such problems can substantially increase the amount of data required for a given accuracy of prediction (Meng, 2018).

are receiving information and/or entertainment in exchange for data and/or attention. If we attend to content co-presented with advertisements, we afford the presenters an opportunity to influence our behaviour. The same is true for less conventional forms of nudging, for example the postulated political interventions mentioned in Sec. 3.3. However, these exchanges are only denominated (if at all) in aggregate, and only when the the corporation providing such service is valuated. Much of the data is even collected on speculation, it may be of no or little value until an innovative use is conceived years later.

Our increasing failure to be able to denominate revenue at the traditional point—income, or exchange—may be another cause for increasing wealth inequality, as less of the economy is recognised, taxed, and redistributed. An obvious solution would be to tax wealth directly—for example, the market value of a corporation—rather than income. The information age may make it easier to track the distribution of wealth, making this strategy more viable than it has been in the past, particularly relative to the challenges of tracking income, if the latter challenges are indeed increasing as I described. However, it is inadequate if that wealth is then taxed only in the country (often a tax haven) in which the corporation is formally incorporated. Given that we can see the transnational transfer of data and engagement with services, we should in theory be able to disseminate redistribution in proportion to the extent and value of data derived. Enforcing such a system transnationally would require substantial innovations, since ordinarily taxation is run by government, and there is almost definitionally no transnational government. There are however international treaties and organised economic areas. Large countries or coordinated economies such as the European Economic Area may be able to demand equitable redistribution for their citizens in exchange for the privilege of access to those citizens. China has successfully demonstrated that such access is not necessarily a given, and indeed blocking access can facilitiate the development of local competition. Similar strategies are being used by American cities and states against platforms such as Uber and AirBnB.

Taxation of ICT wealth leads me to a proposed distortion of law that is particularly dangerous. In 2016 the European Parliament proposed that AI or robotics might be reasonably taxed as 'e-persons'. This is a terrible idea (Bryson et al., 2017). It would allow corporations to automate part of their business process, then break off that piece in such a way as to limit their liabilities for both taxes and legal damages.

The idea of taxing robots has populist appeal for two reasons. First, it seems basic common sense that if robots are "taking our jobs" they should also "pay taxes" and thus support "us" via the welfare state. Second, many people find appealing the idea that we might extend human life—or something more essential about humanity than life—synthetically via AI and/or robotics. Unfortunately, both of these ideas are deeply incoherent, resting on ignorance about the nature of intelligence.

As described in Sec 3.1 earlier, both of these appeals assume that *intelligent* means in part *humanlike*. While there is no question that the word has been used that way culturally, by the definitions presented in Sec 2 it is clearly completely false. To address the second concern first, the values, motivations, even the aesthetics of an enculturated ape cannot be meaningfully shared with a device that shares nothing of our embodied physical ('phenomenological') experience (Bryson, 2008; Claxton, 2015; Dennett, 2017). Nothing we build from metal and silicon will ever share our phenomenology as much as a rat or cow, and few see cows or rats as viable vessels of our posterity.

Further, the idea that substantiating a human mind in digital technology—even were that possible—would make it immortal or even increase its lifespan is ludicrous. Digital formats have a mean lifetime of no more than five years (Lawrence et al., 2000; Marshall et al., 2006). The fallacy here is again to mistake computation for a form of mathematics. While mathematics really is pure, eternal, and certain, that is because it is also not real—it is not manifest in the physical world and cannot take actions. Computation in contrast is real. As described earlier, computation takes time, space, and energy (Sipser, 2005). Space is needed for storing state (memory), and there is no permanent way to achieve such storage (Krauss and Starkman, 2000).

To return to the seemingly more practical idea of taxing AI entities, this again overlooks their lack of humanity. In particular, AI is not countable as humans are countable. This criticism holds also for Bill Gates' support of taxing robots, even though he did not support legal personality (author pool, 2017). There is no equivalent of "horsepower" to measure the number of humans replaced by an algorithm. As just mentioned, in the face of accelerating innovation we can no longer keep track of the value even of transactions including human participants. When a new technology is brought in, we might briefly see how many humans are made redundant, but even this seems to reflect more the current economy than the actual value of labour replaced (Autor, 2015; Ford, 2015). When times are good, a company will retain and

retrain experienced employees; when times are bad corporations will take the excuse to reduce headcount. Even if the initial shift in employment were indicative of initial "personpower" replaced, technologies quickly change the economies into which they are inserted, and the values of human labour too rapidly change.

It is essential to remember that artefacts are by definition designed. Within the limits of the laws of physics and computation, we have complete authorship over AI and robotics. This means that developers will be able to evade tax codes in ways inconceivable to legislators used to value based on human labour. The process of decomposing a corporation into automated 'e-persons' would enormously magnify the present problems of the over-extension of legal personhood such as the shell corporations used for money laundering. The already restricted sense in which it is sensible to consider corporations to be legal persons would be fully dissolved if there are no humans employed by the synthetic entity (Solaiman, 2016; Bryson et al., 2017).

4. The Next Ten Years: Remediations and Futures

To stress again as at the beginning of Sec 3, AI has been and is an incredible force of both economic growth and individual empowerment. We are with it able to know, learn, discover, and do things that would have been inconceivable even 50 years ago. We can walk into a strange city not knowing the language yet find our way and communicate. We can take advantage of education provided by the world's best universities in our own homes, even if we are leading a low-wage existence in a developing economy (Breslow et al., 2013). Even in the developing world, we can use the village smart phone to check the fair prices of various crops, and other useful information like weather predictions, so even subsistence farmers are being drawn out of extreme poverty by ICT. The incredible pace of completion of the Human Genome Project is just one example of how humanity as a whole can benefit from this technology (Adams et al., 1991; Schena et al., 1996).

Nevertheless, the concerns highlighted above need to be addressed. I will here make suggestions about each, beginning with the most recently presented. I will be brief here, since as usual knowledge of solutions only follows from identification of problems, and the identifications above are not yet agreed but only proposed. In addition, some means for redressing these

issues have already been suggested, but I go into further and different detail here.

4.1. Employment and Social Stability

I have already in Sec 3.4 dismissed the idea that making AI legal persons would address the problems of employment disruption or wealth inequality we are currently experiencing. In fact e-personhood would almost certainly *increase* inequality by shielding companies and wealthy individuals from liability, at the cost of the ordinary person. We have good evidence now that wealthy individual donors can lead politicians to eccentric, extremist position taking (Barber, 2016) which can lead to disastrous results when coupled with increasing pressure for political polarisation and identity politics. It is also important to realise that not every extremely wealthy individual necessarily reveals the level of their wealth publicly.

In democracies, another correlate of periods of high inequality and high polarisation is very close elections, even where candidates might otherwise not seem evenly matched. This of course opens the door to (or at leat reduces the cost of) manipulation of elections, including by external powers. Person (2018) suggests weak countries may be practicing 'subtractive balancing' against stronger ones, by disrupting elections and through them governance abilities and therefore autonomy, in an effort to reduce power differentials in the favour of the weaker nation. If individuals or coalitions of individuals are sufficiently wealthy to reduce the efficacy of governments, then states also lose their autonomy, including the stability of their borders.

War, anarchy, and their associated instability is not a situation anyone should really want be in, though those who presently profit from illegal activity might think otherwise. Everyone benefits from sufficient stability to plan businesses and families. The advent of transnational corporations has been accompanied by a substantial increase in the number and power of other transnational organisations. These may be welcome if they help coordinate cooperation on transnational interests, but it is important to realise that geography will always be a substantial determiner of many matters of government. How well your neighbour's house is protected from fire, whether their children are vaccinated or well educated, will always affect your quality of life. Fresh water, sewage, clean air, protection from natural disasters, protection from invasion, individual security, access to transport options—local and national governments will continue to play an extremely important role

in the indefinite future even if in some functions are offloaded to corporations or transnational governments. As such, they need to be adequately resourced.

I recommended in Sec. 3.4 that one possible solution to the impact on ICT on inequality is to shift priority from documenting and taxing income to documenting and taxing wealth. The biggest problem with this suggestion may be that it requires redistribution to occur internationally, not just internationally, because the richest corporations per Internet[9] are in only one country, though certainly for those outside China—and increasingly for those inside—their wealth derives from global activity. Handling this situation will require significant policy innovations. Fortunately, it is in the interest of nearly all stakeholders, including leading corporations, to avoid war and other destructive social and economic instability. The World Wars and financial crises of the Twentieth Century showed that this was especially true the extremely affluent, who at least economically have the most to lose (Milanovic, 2016; Scheidel, 2017), though of course do not often lose their lives.

I particularly admire the flexible solutions to economic hardship that Germany displayed during the recent recession, where it was possible for corporations to *partially* lay off employees, who then received *partial* welfare and free time (Eichhorst and Marx, 2011, p. 80). This allowed individuals to retrain while maintaining for a prolonged period a standard of living close to their individual norms; it also allowed companies to retain valued employees while they pivoted business direction or just searched for liquidity. This kind of flexibility should be encouraged, with both governments and individuals retaining economic capacity to support themselves through periods of crisis. In fact, sufficient flexibility may prevent periods of high change from being periods of crisis.

If we can reduce inequality, I believe the problems of employment will also reduce, despite any increase in the pace of change. We are a fabulously wealthy society, and can afford to support individuals at least partially as they retrain. We are also fantastically innovative. If money is circulating in communities, then individuals will find ways to employ each other, and to

[9]The world is effectively split into two Internets, one inside and one outside the Great Firewall (Ensafi et al., 2015). Both sides similarly contain a small number of extremely successful companies operating in the digital economy (Yiu, 2016; Dolata, 2017).

perform services for each other (Hunter et al., 2001; Autor, 2015). Again, this may already be happening, and could account for the decreased rate of change some authors claim to detect in society (e.g. Cowen, 2011). A great number of individuals may continue finding avenues of self and (where successful) other employment in producing services within their own communities, from the social such as teaching, policing, journalism, and family servies, to the aesthetic such as personal, home, and garden decoration, and the provision of food, music, sports, and other communal acts.

The decision about whether such people are able to live good enough lives that they can benefit from the advantages of their society is a matter of economic policy. We would want any family to be able for example to afford a week's holiday in the nearest large city, or to have their children experience social mobility, for example getting into the top universities in their chosen area purely based on merit. Of course we expect in this century universal and free access to healthcare, and primary and secondary education. People should be able to live with their families but also not need to commute for enormous portions of their day, this requires both distributed employment opportunities and excellent, scalable (and therefore probably public) transportation infrastructure.

The level of investment in such infrastructure depends in part on the investment both public and private in taxation, and also on how such wealth is spent. Historically we have in some periods spent a great deal on the destruction of others' infrastructure and repair of ones own due to warfare. Now even if we avoid open ballistic warfare, we must face the necessity of abandoning old infrastructure that is no longer viable due to climate change, and investing in other locations. Of course, this offers a substantial opportunity for redistribution, particularly into some currently economically depressed cities, as was shown by Roosevelt's New Deal, which substantially reduced inequality in the USA well before the second world war (McCarty et al., 2016; Wright, 1974).

I am with those who do not believe the universal basic income is a great mechanism of redistribution, for several reasons. First, many hope to fund it by cutting public services, but these may well be increasingly needed as increasing numbers of people cannot deal with the technical and economic complexities of a world of accelerating change. Second, I have seen far too many standing safely in the middle of road telling television cameras that "the government has never done anything for me", ignorant of massive investment in their education, security, and infrastructure. I think a basic income

would easily become as invisible and taken for granted as trash collection and emergency services apparently are.

But most importantly, I would prefer redistribution to reenforce the importance of local civic communities, that is to circulate through employment, whether direct or as freelancers and customers. AI and ICT make it easy to bond with people from all over the world, or indeed with entertaining fantasies employing AI technology that are not actually human. But our neighbours' well being has enormous impacts on our own and are in many senses shared, through the quality of water, air, education, fire and other emergency services, and of course personal security. The best neighbourhoods are connected through knowledge and personal concern, that is localised friendships.

One effective mechanism of increasing redistribution is just the increase of minimum wages (Lee, 1999; Schmitt, 2013). Even if this is only done for government employees, it has knock-on effects for the rest of employers as they compete for the best people, and of course also also gives the advantage of having better motivation for good workers to contribute to society through civil service. Although this mechanism has been attacked for a wide variety of reasons (e.g. Meyer, 2016), the evidence seems fairly good for positive impacts overall.

4.2. Privacy, Liberty, and Innovation

Stepping back to the coupled problems of privacy and individual autonomy, we hit an area for which predictions are more difficult or at least more diverse. It is clear that the era of privacy through obscurity is over, as we now have more information and more means to filter and understand information than ever before, and this is unlikely to be changed by anything short of a global disaster eliminating our digital capacity. Nevertheless, we have long been in the situation of inhabiting spaces where our governments and neighbours could in theory take our private property from us, but seldom do except by contracted agreement such as taxation (Christians, 2009). Can we arrive at a similar level of control over our personal data? Can we have effective privacy and autonomy in the information era? If not, what would be the consequences?

First it should be said that any approach to defending personal data and protecting citizens from being predicted, manipulated, or outright controlled via their personal data requires strong encryption and cybersecurity—*without* back doors. Every back door in cybersecurity has been exploited by bad actors (Abelson et al., 2015). Weak cybersecurity should be viewed as a

significant risk to the AI and digital economy, particularly the Internet of Things (IoT). If intelligent or even just connected devices cannot be trusted, they will and should not be welcome in homes or workplaces (Singh et al., 2016; Weber, 2010).

Many thinkers on the topic of technologically mediated privacy have suggested that data about a person should be seen not as an asset of the person but as *part* of that person — an extension of an individual's identity. As such, personal data cannot be owned by anyone but the person to whom it refers; any other use is by lease or contract which cannot be extended or sold onwards without consent (Crabtree and Mortier, 2015; Gates and Matthews, 2014). This would make personal data more like your person; if it and therefore you are violated, you should have recourse to the law. There are a variety of legal and technological innovations being developed in order to pursue this model, however given both the ease of access to data and the difficulty of proving such access, data may be far more difficult to defend than physical personal property (Rosner, 2014; Jentzsch, 2014). Fortunately, at least some governments have made it part of their job to defend the data interests of their citizens (e.g. the GDPR Albrecht, 2016; Danezis et al., 2014). This is for excellent reasons, since as described above there are both political and economic consequences of foreign extraction of data wealth and manipulation of individual political preferences and other behaviour based on those individual's social media profiles.

The best situated entities to defend our privacy are governments, presumably through class-action lawsuits of at least the most egregious examples of violation of personal data. Note that such courses of action may require major innovations of international law or treaties, since some of the most prominent allegations of manipulation involve electoral outcomes for entire countries. For example, the UK's Brexit vote has in the first two years since the referendum (and before any actual exit of the EU) cost the country £23bn in lost tax revenue, or £44mn a week (Morales, 2018). As mentioned earlier the Brexit vote is alleged to have been influenced by known AI algorithms, which have been shown to have been funded through foreign investment (ICO, 2018). Ironically, achieving compensation for such damage would almost certainly require international collaboration.

Unfortunately, governments do not always have their citizens' interests at heart, or at least, not always all of their citizens' interests. Indeed, globally in the Twentieth Century, one was far more likely to be killed by one's own government than by any foreign actor (Valentino, 2004). More recently,

China has been using the surveillance system that was supposed to keep its citizens safe to destroy the lives and families of over a million of its citizens by placing them in reeducation camps for the crime of even casually expressing their Muslim identity (Human Rights Watch, 2018; Editor, 2018). More generally, if governments fear whistle blowing, dissent, or even just shirk the responsibility for guaranteeing dignity and flourishing for all those in their territory, then they can and often will suppress and even kill those individuals. It is extremely dangerous when a government views governing a group of people within its borders as more cost or trouble than their collective potential value for labour, security, and innovation is worth. Aggravating this grave threat, we have both the promise and the hype of intelligent automation as a new, fully ownable and controllable source of both labour and innovation. The inflated discourse around AI increases the risk that a government will (mis)assesses the value of human lives to be lower than the perceived costs of maintaining those lives.

We cannot know the sure outcome of the current trajectories, but where any sort of suppression is to be exercised, we can readily expect that AI and ICT will be the means for monitoring and predicting potential trouble makers. China is alleged to be using face-recognition capacities not only to identify individuals, but to identify their moods and attentional states both in reeducation camps and in ordinary schools. Students and perhaps teachers can be penalised if students do not pay attention, and prisoners can be punished if they do not appear happy to comply with their (re)education. ICT systems able to detect and inform teachers to adjust lectures and material towards students' attention and comprehension are also being pitched for classrooms in the West, and are core to personalised AI instruction outside of conventional classrooms. Presumably similar systems are also being developed and probably applied for other sorts of work (e.g. Levy, 2015).

If we allow such trends to carry on, we can expect societies that are safer—or at least more peaceful on the streets—more homogenous, less innovative, and less diverse. More people have the means and economic wherewithal to move between countries now than ever before, so we might hope that countries that truly produce the best quality of life including governance and individual protection will be attractors to those who care about personal liberty. We may also hope that with the combined power of those immigrants and their extant citizens' labour and innovation, these countries may come to be able to protect not only themselves but others. We've already seen the EU do such protection by setting standards of AI ethics such as the

GDPR, and of course the United Nations is working with instruments such as the Paris Agreement to protect us all from from climate change. In such well-governed and flourishing societies, we would expect to see perhaps an increase rather than a decrease in present levels of liberty, as we come to recognise the problems arising from the surveillance we already practice, e.g. in micromanaging our children's personal time (Lee et al., 2010; Bryson, 2015).

Unfortunately for this optimistic vision of pools of wellbeing spreading from well-governed countries, in practice technology is increasingly being used or being threatened to be used for blocking any cross-border migration except by the most elite (Miller, 2017). Besides genocide and mass killings, another historic trend often observed in wars and political revolutions (e.g. Nazi occupied Poland, cold war Czechoslovakia, The Iranian Revolution, Stalin's USSR, Cambodia under the Khmer Rouge, China's Cultural Revolution, present-day Saudi Arabia) is the displacement or even execution of not only dissidents but all and any intelligentsia. The effort to maintain control is often seen as requiring the elimination of any potential innovative leadership, even though precisely such leadership may be necessary to keep people healthy and therefore civil society stable (King and Zeng, 2001), not to mention maintaining the technological progress necessary to stay abreast in any arms race (Yan, 2006). Such movements tend to fail only after protracted suffering, and often only after having persisted long enough to make clear the damage of their policies to the country's international competitiveness. AI makes the identification and isolation of any such targeted group—or even individuals with target attitudes—spectacularly easy. Only if we are able to also innovate protections against corrupt, selfish, or otherwise dangerous governance can we protect ourselves from losing the diversity and liberty of our societies, and therefore the security of us all.

Again, the capacity for AI to be used for good governance leading to fairer and stronger societies is very real and widely being developed. For example, AI is used to reduce financial crime, fraud, and money laundering, protecting individuals, societies, and governments from undue and illegal influence (Ngai et al., 2011). This is sensible and a part of ordinary contractual understandings of the duties of financial service providers and indeed governments. It may also be ethical for citizens to be 'nudged' by their devices or other agencies into behaviours they have consciously chosen and explicitly expressed a desire for, such as exercise or sleep regimes. But it is important to recognise the massively increased threats of both explicit duress and

implicit misleading that accompanies the massive increase of knowledge and therefore power that derive from AI. AI therefore increases the urgency of investment in research and development in the humanities and social sciences, particularly the political sciences and sociology. I therefore now turn to the problem of regulating AI.

4.3. Individual, Corporate, and Regulatory Responsibility for AI

To begin the discussion of responsibility in an age of AI, I want to return briefly to emphasise again the role of design and architectures on AI. Again perhaps because of the misidentification of *intelligent* with *human*, I have sometimes heard even domain experts from influential organisations claim that one or another trait of AI is inevitable. There is no aspect of AI more inevitable than slavery or the hereditary rights of monarchy. Of course, both of these still persist in some places, but despite their economic benefits to those formerly in power, they have been largely eradicated. Similarly, we can regulate at least legal commercial products to mandate safe or at least transparent architectures (Boden et al., 2011). We can require—as again the European Commission recently has—that decisions taken by machines be traceable and explainable (Goodman and Flaxman, 2016).

Maintaining human accountability for AI systems does not have to mean that we must (or can) account for the value of every weight in a machine-learned neural network, or the impact of every individual instance of data used in training. Not only is this impractical, it is not the standard or means by which we currently hold organisations accountable. A company is not responsible for the synaptic organisation of its accounts' brains, it is responsible for the state of its accounts. Introducing AI into a corporate or governance process actually changes little with respect to responsibility. We still need to be able to characterise our systems well enough to recognise whether they are behaving as intended (Liu et al., 2017). This is doable, and it should be encouraged (Bryson and Theodorou, 2019).

Encouraging responsibility entails ensuring we continue maintaining accountability (Santoni de Sio and van den Hoven, 2018). One simple way to do this is to educate governments and prosecutors that software systems have much the same liability issues as any other manufactured artefact—if they are misused, it is the fault of the owner; if they cause harm when being appropriately used, they are at fault and the manufacturer is likely liable unless they can prove due diligence and exceptional circumstance. The mere fact that part of the system is autonomous does not alter this fact, just as

a bank can be held accountable for errors generated by its accountants or customers where they bank's systems should have caught or constrained such errors. There are certainly challenges here, particularly because so many applications of AI technology are in transnational contexts, but organisations such as the EU, UN, and OECD are looking to be able to coordinate the efforts of nations to protect their citizens.

Of course AI systems are not exactly like more deterministic systems, but exaggerating the consequences of those differences creates problems. Bad ideas can be hidden by the "smoke and mirrors" of the confusion generated by AI around identity and moral agency (Bryson, 2018). One concerning trend in AI governance that concerning is the trend for *value alignment* as a solution to difficult questions of science generally, and AI ethics in particular. The idea is that we should ensure that society leads and approves of where science or technology go (Soares and Fallenstein, 2014). This may sound very safe and democratic, but it is perhaps better seen as populist. Speaking first to science: science is a principle mechanism enabling society to accurately *perceive* its context. In contrast, *governance* is how a society chooses between potential actions. Popular sentiment cannot determine what is true about nature; it can only determine what policies are easiest to deploy. To limit a society's capacity to perceive to only the things it wants to know would be to blind that society (Caliskan et al., 2017, see the final discussion). Similarly, the outcomes of policy are highly influenced by public sentiment, but certainly not determined by it. Asking the public what it wants from AI is like asking them which science fiction film they would most like to see realised—there is no guarantee they will choose one that is feasible, let alone truly desirable in protracted detail. While the public must through government determine its economic and political priorities, actual progress is almost never achieved by referendum. Rather, governance almost always comes down to informed negotiations between a limited number of expert negotiators, supported by a larger but still limited number of domain experts.

Even given the vast resources available through exploiting computation and AI, it is likely that human negotiators will always be the best determiners of policy. This is partly because we as citizens can identify with human representatives, thus establishing trust and investment in the negotiated outcomes. But more importantly, human representatives can be held to account and persuaded in ways that AI never can be. We cannot intentionally design systems to be as centred on social outcomes as human or indeed any

social animal's intelligence has evolved to be. We cannot do this by design, because design by its nature is a decomposable process, whereas evolution has repeatedly discovered that concern for social standing must be an inextricable part of an individual's intelligence for a species reliant on social strategies for its survival. Thus are entire system of just relies on dissuasion to do with isolation, loss of power or social standing. We cannot apply such standards of justice to machines we design, and we cannot trace accountability through machines we do not carefully design (Bryson et al., 2017; Bryson and Theodorou, 2019).

Finally, some have expressed concern that it is impossible to maintain regulation in the face of AI because of the rapid rate of change AI entails. It is true that individual humans have limits in their capacities, including on how quickly we can respond. Similarly, legislation can only be written at a particular pace. Latencies are in fact deliberately built into the legislative process to ensure the pace of change is not too high for business and personal planning (Holmes, 1988; Cowen, 1992; Ginsburg, 2005; Roithmayr et al., 2015). Therefore legislation alone cannot be expected to keep up with the accelerating pace of change brought on by AI and ICT. I have previously suggested that one mechanism for forging sensible policy is to have domain experts working through professional organisations describe systems of standards (Bryson and Winfield, 2017). The role of government then is reduced to monitoring those efforts and lending enforcement to their outcomes. The arguments I have made above (and in Bryson and Theodorou, 2019) might be seen as a generalisation of this principle. Here we are saying that we do not need to change legislation at all, simply hold organisations that build or exploit AI to account for the consequences for their systems' actions by the ordinary and established means of tracing accountability. It is then these organisations who will need to do the innovation on accountability in lock step with their other innovation, so that they can demonstrate that they have always followed due diligence with their systems.

5. Conclusion

Artificial intelligence is already changing society at a faster pace than we realise, but at the same time it is not as novel or unique in human experience as we are often lead to imagine. Other artifactual entities, such as language and writing, corporations and governments, telecommunication and oil, have previously extended our capacities, altered our economies, and disrupted our

social order—generally though not universally for the better. The evidence that we are on average better off for our progress is ironically perhaps the greatest threat we currently need to master: sustainable living and reversing the collapse of biodiversity.

Nevertheless AI—and ICT more generally—may well require radical innovations in the way we govern, and particularly in the way we raise revenue for redistribution. We are faced with transnational wealth transfers through business innovations that have outstripped our capacity to measure or even identify the level of income generated. Further, this new currency of unknowable value is often personal data, and personal data gives those who hold it the immense power of prediction over the individuals it references.

But beyond the economic and governance challenges, we need to remember that AI first and foremost extends and enhances what it means to be human, and in particular our problem-solving capacities. Given ongoing global challenges such as security and sustainability, such enhancements promise to continue to be of significant benefit, assuming we can establish good mechanisms for their regulation. Through a sensible portfolio of regulatory policies and agencies, we should continue to expand—and also to limit, as appropriate—the scope of potential AI application.

Acknowledgements

I would like to acknowledge my collaborators, particularly Karine Perset for recruiting me to work on these topics for the OECD and for many good conversations, my (recent) PhD students Andreas Theodorou and Rob Wortham, Alan Winfield with whom some of this content has been reworked and extended to consider the role of professional societies (see Bryson and Winfield, 2017), Karen Croxson of the UK's Financial Conduct Authority, and Will Lowe, particularly for feedback on the sections concerning international relations. Thanks to the OECD for permission to reuse some of the material above for academic contexts such as this volume. I also thank the AXA Research Fund for part-funding my work on AI ethics from 2017–2020.

Abelson, H., Anderson, R., Bellovin, S. M., Benaloh, J., Blaze, M., Diffie, W., Gilmore, J., Green, M., Landau, S., Neumann, P. G., Rivest, R. L., Schiller, J. I., Schneier, B., Specter, M. A., and Weitzner, D. J. (2015). Keys under doormats: Mandating insecurity by requiring government access to all data and communications. *Journal of Cybersecurity*, 1(1):69.

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991). Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science*, 252(5013):1651–1656.

Aker, J. C. and Mbiti, I. M. (2010). Mobile phones and economic development in africa. *Journal of Economic Perspectives*, 24(3):207–232.

Albrecht, J. P. (2016). How the GDPR will change the world. *European Data Protection Law Review*, 2(3).

Armstrong, H. and Read, R. (1995). Western European micro-states and EU autonomous regions: The advantages of size and sovereignty. *World Development*, 23(7):1229–1245.

Assael, Y. M., Shillingford, B., Whiteson, S., and de Freitas, N. (2016). LipNet: Sentence-level lipreading. *arXiv preprint arXiv:1611.01599*.

author pool (2017). I, taxpayer: Why taxing robots is not a good idea — Bill Gates's proposal is revealing about the challenge automation poses. *The Economist*. print edition.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *The Journal of Economic Perspectives*, 29(3):3–30.

Bandyopadhyay, A. and Hazra, A. (2017). A comparative study of classifier performance on spatial and temporal features of handwritten behavioural data. In Basu, A., Das, S., Horain, P., and Bhattacharya, S., editors, *Intelligent Human Computer Interaction: 8th International Conference, IHCI 2016, Pilani, India, December 12-13, 2016, Proceedings*, pages 111–121. Springer International Publishing, Cham.

Barber, M. J. (2016). Ideological donors, contribution limits, and the polarization of american legislatures. *The Journal of Politics*, 78(1):296–310.

Barnosky, A. D. (2008). Megafauna biomass tradeoff as a driver of quaternary and future extinctions. *Proceedings of the National Academy of Sciences*, 105(Supplement 1):11543–11548.

Barrett, D. P., Barbu, A., Siddharth, N., and Siskind, J. M. (2016). Saying what you're looking for: Linguistics meets video search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):2069–2081.

Barrows, E. M. (2000). *Animal behavior desk reference: a dictionary of animal behavior, ecology, and evolution*. CRC press.

Bhorat, H., Naidoo, K., and Pillay, K. (2016). Growth, poverty and inequality interactions in Africa: An overview of key issues. Africa Policy Notes 2016-02, United Nations Development Programme, New York, NY.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, London.

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorell, T., Wallis, M., Whitby, B., and Winfield, A. (2011). Principles of robotics. The United Kingdom's Engineering and Physical Sciences Research Council (EPSRC).

Bostrom, N. (2012). The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. *Minds and Machines*, 22(2):71–85.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., and Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, 8:13–25.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hÉigeartaigh, S. O., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., and Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. Technical report, Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, and OpenAI. https://maliciousaireport.com/.

Brundage, M. and Bryson, J. J. (2017). Smart policies for artificial intelligence. in preparation, available as arXiv:1608.08196.

Brynjolfsson, E., Rock, D., and Syverson, C. (2017). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In *Economics of Artificial Intelligence*. University of Chicago Press.

Bryson, J. J. (2003). The Behavior-Oriented Design of modular agent intelligence. In Kowalszyk, R., Müller, J. P., Tianfield, H., and Unland, R., editors, *Agent Technologies, Infrastructures, Tools, and Applications for e-Services*, pages 61–76. Springer, Berlin.

Bryson, J. J. (2008). Embodiment versus memetics. *Mind & Society*, 7(1):77–94.

Bryson, J. J. (2015). Artificial intelligence and pro-social behaviour. In Misselhorn, C., editor, *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, volume 122 of *Philosophical Studies*, pages 281–306. Springer, Berlin.

Bryson, J. J. (2018). Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology*, 20(1):15–26.

Bryson, J. J., Diamantis, M. E., and Grant, T. D. (2017). Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law*, 25(3):273–291.

Bryson, J. J. and Theodorou, A. (2019). How society can maintain human-centric artificial intelligence. In Toivonen-Noro, M. and Saari, E., editors, *Human-Centered Digitalization and Services*. Springer.

Bryson, J. J. and Winfield, A. F. T. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, 50(5):116–119.

Bullinaria, J. A. and Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, 39(3):510–526.

Bullough, O. (2018). The rise of kleptocracy: The dark side of globalization. *Journal of Democracy*, 29(1):25–38.

Cadwalladr, C. (2017a). Revealed: How US billionaire helped to back Brexit — Robert Mercer, who bankrolled Donald Trump, played key role with 'sinister' advice on using Facebook data. *The Observer*.

Cadwalladr, C. (2017b). Robert Mercer: The big data billionaire waging war on mainstream media. *The Observer*.

Calinon, S., D'halluin, F., Sauser, E. L., Caldwell, D. G., and Billard, A. G. (2010). Learning and reproduction of gestures by imitation. *IEEE Robotics & Automation Magazine*, 17(2):44–54.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Chessell, M. and Smith, H. C. (2013). *Patterns of information management*. Pearson Education.

Christians, A. (2009). Sovereignty, taxation and social contract. *Minn. J. Int'l L.*, 18:99.

Chung, J. S. and Zisserman, A. (2017). Lip reading in the wild. In Lai, S.-H., Lepetit, V., Nishino, K., and Sato, Y., editors, *Computer Vision — ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part II*, pages 87–103. Springer International Publishing, Cham.

Claxton, G. (2015). *Intelligence in the flesh: Why your mind needs your body much more than it thinks*. Yale University Press.

Cooke, M. (1999). A space of one's own: Autonomy, privacy, liberty. *Philosophy & Social Criticism*, 25(1):22–53.

Cowen, T. (1992). Law as a public good: The economics of anarchy. *Economics and Philosophy*, 8(02):249–267.

Cowen, T. (2011). *The great stagnation: How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better*. Penguin.

Crabtree, A. and Mortier, R. (2015). Human data interaction: Historical lessons from social studies and CSCW. In Boulus-Rødje, N., Ellingsen, G.,

Bratteteig, T., Aanestad, M., and Bjørn, P., editors, *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*, pages 3–21. Springer International Publishing, Cham.

Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Metayer, D. L., Tirtea, R., and Schiffner, S. (2014). Privacy and data protection by design-from policy to engineering. Technical report, European Union Agency for Netowrk and Information Security (ENISA), Heraklion, Greece.

Deng, J., Xu, X., Zhang, Z., Frühholz, S., and Schuller, B. (2017). Universum autoencoder-based domain adaptation for speech emotion recognition. *IEEE Signal Processing Letters*, 24(4):500–504.

Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. WW Norton & Company.

Dennett, D. C. (2017). *From Bacteria to Bach and Back*. Allen Lane.

Devos, T. and Banaji, M. R. (2003). Implicit self and identity. *Annals of the New York Academy of Sciences*, 1001(1):177–211.

Dolata, U. (2017). Apple, Amazon, Google, Facebook, Microsoft: Market concentration-competition-innovation strategies. SOI Discussion Paper 2017-01, Stuttgarter Beiträge zur Organisations-und Innovationsforschung.

Editor (2018). China's Orwellian tools of high-tech repression. *The Washington Post*.

Eichhorst, W. and Marx, P. (2011). Reforming German labour market institutions: A dual path to flexibility. *Journal of European Social Policy*, 21(1):73–87.

Ensafi, R., Winter, P., Mueen, A., and Crandall, J. R. (2015). Analyzing the Great Firewall of China over space and time. *Proceedings on privacy enhancing technologies*, 2015(1):61–76.

Erickson, B. J., Korfiatis, P., Akkus, Z., Kline, T., and Philbrick, K. (2017). Toolkits and libraries for deep learning. *Journal of Digital Imaging*, pages 1–6.

Eyben, F., Weninger, F., Lehment, N., Schuller, B., and Rigoll, G. (2013). Affective video retrieval: Violence detection in Hollywood movies by large-scale segmental feature extraction. *PLoS ONE*, 8(12):e78506.

Ford, M. (2015). *Rise of the Robots: Technology and the Threat of a Jobless Future*. Oneworld, London.

Frischmann, B. M. and Selinger, E. (2016). Engineering humans with contracts. Research Paper 493, Cardozo Legal Studies. Available at SSRN: https://ssrn.com/abstract=2834011.

Gabriel, I. (2017). Effective altruism and its critics. *Journal of Applied Philosophy*, 34(4):457–473.

Gates, C. and Matthews, P. (2014). Data is the new currency. In *Proceedings of the 2014 New Security Paradigms Workshop*, NSPW '14, pages 105–116, New York, NY, USA. ACM.

Gilbert, S. F., Sapp, J., and Tauber, A. I. (2012). A symbiotic view of life: We have never been individuals. *The Quarterly Review of Biology*, 87(4):325–341. PMID: 23397797.

Ginsburg, T. (2005). Locking in democracy: Constitutions, commitment, and international law. *NYUJ Int'l. L. & Pol.*, 38:707.

Goodman, B. and Flaxman, S. (2016). EU regulations on algorithmic decision-making and a "right to explanation". In Kim, B., Malioutov, D. M., and Varshney, K. R., editors, *ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, pages 26–30, New York, NY.

Grama, A. (2003). *Introduction to parallel computing*. Pearson Education.

Griffin, H. J., Aung, M. S. H., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., and Bianchi-Berthouze, N. (2013). Laughter type recognition from whole body motion. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 349–355, Geneva, CH.

Haberl, H., Erb, K. H., Krausmann, F., Gaube, V., Bondeau, A., Plutzar, C., Gingrich, S., Lucht, W., and Fischer-Kowalski, M. (2007). Quantifying and mapping the human appropriation of net primary production in Earth's

terrestrial ecosystems. *Proceedings of the National Academy of Sciences*, 104(31):12942–12947.

Haines, T. S. F., Mac Aodha, O., and Brostow, G. J. (2016). My text in your handwriting. *ACM Trans. Graph.*, 35(3):26:1–26:18.

Hanahan, D. and Weinberg, R. (2011). Hallmarks of cancer: The next generation. *Cell*, 144(5):646–674.

Hancock, J. T., Curry, L. E., Goorha, S., and Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45(1):1–23.

Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868):1362–1367.

Hofmann, M., Geiger, J., Bachmann, S., Schuller, B., and Rigoll, G. (2014). The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, 25(1):195–206.

Holmes, S. (1988). Precommitment and the paradox of democracy. *Constitutionalism and democracy*, 195(195):199–221.

Hsu, F.-h. (2002). *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*. Princeton University Press.

Huang, B., Li, M., De Souza, R. L., Bryson, J. J., and Billard, A. (2016). A modular approach to learning manipulation strategies from human demonstration. *Autonomous Robots*, 40(5):903–927.

Human Rights Watch (2018). "eradicating ideological viruses": China's campaign of repression against xinjiang's muslims. Technical report, Human Rights Watch.

Hunter, L. W., Bernhardt, A., Hughes, K. L., and Skuratowicz, E. (2001). It's not just the ATMs: Technology, firm strategies, jobs, and earnings in retail banking. *Industrial & Labor Relations Review*, 54(2):402–424.

ICO (2018). Investigation into the use of data analytics in political campaigns: Investigation update. Technical report, Information Commissioner's Office, United Kingdom.

ICRC (2018). Ethics and autonomous weapon systems: An ethical basis for human control? Technical report, International Committee of the Red Cross, Geneva.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, not ideology: Social identity perspective on polarization. *Public Opinion Quarterly*, 76(3):405.

Jentzsch, N. (2014). Secondary use of personal data: A welfare analysis. *European Journal of Law and Economics*, pages 1–28.

Jia, S., Lansdall-Welfare, T., and Cristianini, N. (2016). Time series analysis of garment distributions via street webcam. In Campilho, A. and Karray, F., editors, *Image Analysis and Recognition: 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel, Póvoa de Varzim, Portugal, July 13-15, 2016, Proceedings*, pages 765–773. Springer International Publishing, Cham.

Jordan, J. J., Hoffman, M., Nowak, M. A., and Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences*.

King, G. and Zeng, L. (2001). Improving forecasts of state failure. *World Politics*, 53:623–658.

Kleinsmith, A. and Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *Affective Computing, IEEE Transactions on*, 4(1):15–33.

Krauss, L. M. and Starkman, G. D. (2000). Life, the universe, and nothing: Life and death in an ever-expanding universe. *The Astrophysical Journal*, 531(1):22.

Laland, K. N., Odling-Smee, J., and Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences*, 23(1):131–146.

Lamba, S. and Mace, R. (2012). The evolution of fairness: Explaining variation in bargaining behaviour. *Proceedings of the Royal Society B: Biological Sciences*.

Landau, J.-P. (2016). Populism and debt: Is Europe different from the U.S.? Talk at the Princeton Woodrow Wilson School, and in preparation.

Lawrence, G. W., Kehoe, W. R., Rieger, O. Y., Walters, W. H., and Kenney, A. R. (2000). Risk management of digital information: A file format investigation. Menlo College Research Paper 93, Council on Library and Information Resources, Washington, D.C.

Lee, D. S. (1999). Wage inequality in the united states during the 1980s: Rising dispersion or falling minimum wage? *The Quarterly Journal of Economics*, 114(3):977–1023.

Lee, E., Macvarish, J., and Bristow, J. (2010). Risk, health and parenting culture. *Health, Risk & Society*, 12(4):293–300.

Levy, K. E. C. (2015). Beating the box: Resistance to surveillance in the united states trucking industry. Dissertation chapter and in prep. manuscript.

Liao, S.-H. (2005). Expert system methodologies and applications: A decade review from 1995 to 2004. *Expert Systems with Applications*, 28(1):93–103.

List, C. and Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press.

Liu, S., Wang, X., Liu, M., and Zhu, J. (2017). Towards better analysis of machine learning models: A visual analytics perspective. *arXiv preprint arXiv:1702.01226*.

Lowe, W. (2001). Towards a theory of semantic space. In *Proceedings of the Twenty-First Annual Meeting of the Cognitive Science Society*, pages 576–581, Edinburgh. Lawrence Erlbaum Associates.

Mace, R. (1998). The co-evolution of human fertility and wealth inheritance strategies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 353(1367):389–397.

Marshall, C. C., Bly, S., and Brun-Cottan, F. (2006). The long term fate of our digital belongings: Toward a service model for personal archives. *Archiving Conference*, 2006(1):25–30.

McCarthy, J. (1983). The little thoughts of thinking machines. *Psychology Today*, 17(12):46–49.

McCarty, N. M., Poole, K. T., and Rosenthal, H. (2016). *Polarized America: The dance of ideology and unequal riches*. MIT Press, Cambridge, MA, second edition.

Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (i): Law of large populations, big data paradox, and the 2016 us presidential election. *The Annals of Applied Statistics*, 12(2):685–726.

Meyer, B. (2016). Learning to love the government: Trade unions and late adoption of the minimum wage. *World Politics*, 68(3):538–575.

Miguel, J. C. and Casado, M. Á. (2016). GAFAnomy (Google, Amazon, Facebook and Apple): The big four and the b-ecosystem. In Gómez-Uranga, M., Zabala-Iturriagagoitia, J. M., and Barrutia, J., editors, *Dynamics of Big Internet Industry Groups and Future Trends: A View from Epigenetic Economics*, pages 127–148. Springer International Publishing, Cham.

Milanovic, B. (2016). Global inequality.

Mill, J. S. (1859). *On Liberty*. John W. Parker and Son, London.

Miller, T. (2017). *Storming the wall: Climate change, migration, and homeland security*. City Lights Books, San Francisco.

Mishel, L. (2012). The wedges between productivity and median compensation growth. Issue Brief 330, Economic Policy Institute, Washington, DC.

Moeslund, T. B. and Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer vision and image understanding*, 81(3):231–268.

Morales, A. (2018). Brexit has already cost the U.K. more than its E.U. budget payments, study shows. *Bloomberg*.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT press.

Newman, E. J., Sanson, M., Miller, E. K., Quigley-McBride, A., Foster, J. L., Bernstein, D. M., and Garry, M. (2014). People with easier to pronounce names promote truthiness of claims. *PLoS ONE*, 9(2):e88671.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3):559–569.

of the President, E. O. (2016). Artificial intelligence, automation, and the economy. Technical report, Executive Office of the US President.

O'Reilly, T. (2017). *WTF? What's the Future and why It's Up to Us*. Random House, New York.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Person, R. (2018). Gray zone tactics as asymmetric balancing. Paper presented at the American Political Science Association Annual Meeting.

Perzanowski, A. and Schultz, J. (2016). *The End of Ownership: Personal Property in the Digital Economy*. MIT Press, Cambridge, MA.

Pinker, S. (2012). *The Better Angels of our Nature: The Decline of Violence in History and Its Causes*. Penguin, London.

Price, G. R. (1972). Fisher's 'fundamental theorem' made clear. *Annals of Human Genetics*, 36(2):129–140.

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. (2016). Generative adversarial text to image synthesis. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 3.

Roithmayr, D., Isakov, A., and Rand, D. (2015). Should law keep pace with society? Relative update rates determine the co-evolution of institutional punishment and citizen contributions to public goods. *Games*, 6(2):124.

Romanes, G. J. (1883). *Animal intelligence*. D. Appleton.

Rosner, G. (2014). Who owns your data? In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 623–628, New York, NY, USA. ACM.

Roughgarden, J., Oishi, M., and Akçay, E. (2006). Reproductive social behavior: Cooperative games to replace sexual selection. *Science*, 311(5763):965–969.

Russell, S. J. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Prentice Hall, Englewood Cliffs, NJ, third edition.

Santoni de Sio, F. and van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, 5:15.

Sartori, G., Orru, G., and Monaro, M. (2016). Detecting deception through kinematic analysis of hand movement. *International Journal of Psychophysiology*, 108:16.

Scheidel, W. (2017). *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century*. Princeton University Press.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences*, 93(20):10614–10619.

Schlager, K. J. (1956). Systems engineering: Key to modern development. *IRE Transactions on Engineering Management*, (3):64–66.

Schmitt, J. (2013). Why does the minimum wage have no discernible effect on employment. Technical Report 22, Center for Economic and Policy Research, Washington, DC.

Schuller, B., Steidl, S., Batliner, A., Hirschberg, J., Burgoon, J. K., Baird, A., Elkins, A., Zhang, Y., Coutinho, E., and Evanini, K. (2016). The interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language. In *Proceedings of Interspeech*.

Selinger, E. and Hartzog, W. (2017). Obscurity and privacy. In Pitt, J. and Shew, A., editors, *Spaces for the Future: A Companion to Philosophy of Technology*. Routledge, New York, NY. in press.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

Singer, P. (2015). *The most good you can do: How effective altruism is changing ideas about living ethically*. Text Publishing, Melbourne.

Singh, J., Pasquier, T., Bacon, J., Ko, H., and Eyers, D. (2016). Twenty security considerations for cloud-supported Internet of Things. *IEEE Internet of Things Journal*, 3(3):269–284.

Sipser, M. (2005). *Introduction to the Theory of Computation*. PWS, Thompson, Boston, MA, second edition.

Smith, B. (2018). Facial recognition technology: The need for public regulation and corporate responsibility. *Microsoft on the Issues*. https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/.

Soares, N. and Fallenstein, B. (2014). Agent foundations for aligning machine intelligence with human interests: A technical research agenda. unpublished white paper, available https://intelligence.org/files/TechnicalAgenda.pdf.

Solaiman, S. M. (2016). Legal personality of robots, corporations, idols and chimpanzees: A quest for legitimacy. *Artificial Intelligence and Law*, pages 1–25.

Stewart, A. J., McCarty, N., and Bryson, J. J. (2018). Explaining parochialism: A causal account for political polarization in changing economic environments. arXiv preprint arXiv:1807.11477.

Stoddart, D. M. (1990). *The Scented Ape: The Biology and Culture of Human Odour*. Cambridge University Press.

Suwajanakorn, S., Seitz, S. M., and Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics (TOG)*, 36(4):95.

Sylwester, K., Mitchell, J., Lowe, W., and Bryson, J. J. (2017). Punishment as aggression: Uses and consequences of costly punishment across populations. in prep.

Technology Council Committee on Technology, N. S. a. (2016). Preparing for the future of artificial intelligence. Technical report, Executive Office of the US President.

Tepperman, J. (2016). *The Fix: How Countries Use Crises to Solve the World's Worst Problems*. Tim Duggan Books, New York.

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., and Nießner, M. (2016). Face2Face: Real-time face capture and reenactment of RGB videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2387–2395.

Touretzky, D. S. (1988). On the proper treatment of thermostats. *Behavioral and Brain Sciences*, 11(1):5556.

Trewavas, A. (2005). Green plants as intelligent organisms. *Trends in plant science*, 10(9):413–419.

Valentino, B. A. (2004). *Final solutions: Mass killing and genocide in the 20th century*. Cornell University Press.

Valstar, M. F. and Pantic, M. (2012). Fully automatic recognition of the temporal phases of facial actions. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 42(1):28–43.

Van Lange, P. A. M., De Bruin, E., Otten, W., and Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence. *Journal of Personality and Social Psychology*, 73(4):733–746.

van Schaik, C., Graber, S., Schuppli, C., and Burkart, J. (2017). The ecology of social learning in animals and its link with intelligence. *The Spanish Journal of Psychology*, 19.

Vincent, J. (2016). Artificial intelligence is going to make it easier than ever to fake images and video. *The Verge*.

Weber, R. H. (2010). Internet of Things: New security and privacy challenges. *Computer Law & Security Review*, 26(1):23–30.

Widrow, B., Rumelhart, D. E., and Lehr, M. A. (1994). Neural networks: Applications in industry, business and science. *Commun. ACM*, 37(3):93–105.

Williams, C. P. (2010). *Explorations in quantum computing*. Springer Science & Business Media.

Winston, P. H. (1984). *Artificial Intelligence*. Addison-Wesley, Boston, MA.

Wolpert, D. H. (1996a). The existence of *a priori* distinctions between learning algorithms. *Neural Computation*, 8(7):1391–1420.

Wolpert, D. H. (1996b). The lack of *a priori* distinctions between learning algorithms. *Neural Computation*, 8(7):1341–1390.

Wright, G. (1974). The political economy of new deal spending: An econometric analysis. *The Review of Economics and Statistics*, 56(1):30–38.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., ukasz Kaiser, Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144.

Yan, X. (2006). The rise of china and its power status. *The Chinese Journal of International Politics*, 1(1):5–33.

Yiu, E. (2016). Alibaba tops Tencent as Asia's biggest company by market value: New-economy companies that owe their revenue to technology and the internet are now more valuable than oil refineries, manufacturers and banks. *South China Morning Post*.

Youyou, W., Kosinski, M., and Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040.